

PATENT ABSTRACTS OF JAPAN

(6)

(11) Publication number : 2004-170552
 (43) Date of publication of application : 17.06.2004

(51) Int.CI.

G10L 15/20
 G10L 11/02
 G10L 15/00
 G10L 15/02
 G10L 15/04
 G10L 21/02

(21) Application number : 2002-334276

(22) Date of filing : 18.11.2002

(71) Applicant : FUJITSU LTD

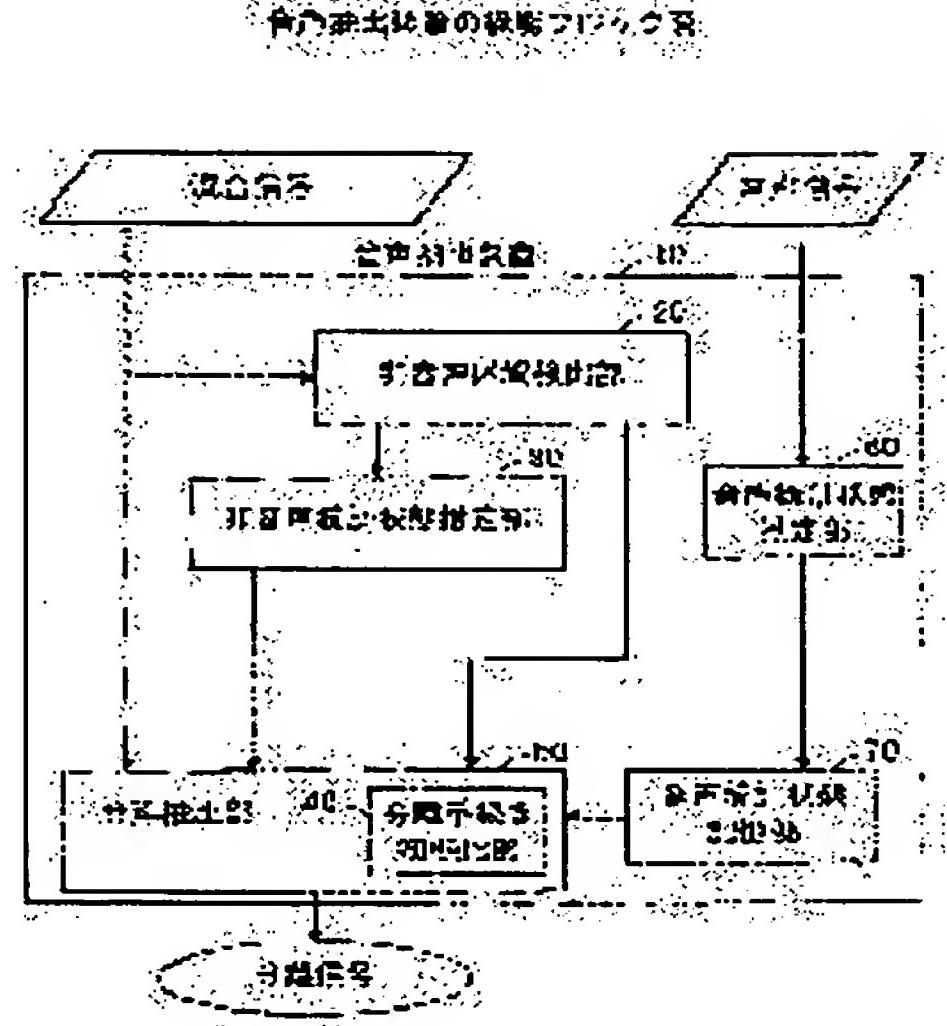
(72) Inventor :
 ENDOU KAORI
 OTA TAKASHI
 SASAKI HITOSHI
 MATSUBARA MITSUYOSHI

(54) SPEECH EXTRACTING APPARATUS

(57) Abstract:

PROBLEM TO BE SOLVED: To provide a speech extracting apparatus that extracts a speech of a specified person from a mixed signal of a speech signal of the specified person and a non-speech signal other than the speech signal while taking a statistical state of a sound source depending upon time into consideration by using a plurality of input signals in which a plurality of sound sources are mixed and the statistical independence of the sound sources included in the input signals.

SOLUTION: A non-speech section detection part 20 detects a non-speech section including no speech from a mixed signal in which a plurality of sound sources are mixed, a non-speech statistical state estimation part 30 estimates a non-speech statistical state by using information on the detected non-speech section, and a speech extraction part 50 extracts at least one speech by sequentially updating the mixed signal according to an initial procedure selected by a separation procedure initialization part 40 by using the statistical independence of the estimated non-speech statistical state and a statistical state, regarding the speech, stored in a speech statistical state storage part 70.



LEGAL STATUS

- [Date of request for examination] 12.05.2005
 [Date of sending the examiner's decision of rejection]
 [Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
 [Date of final disposal for application]
 [Patent number]
 [Date of registration]
 [Number of appeal against examiner's decision of rejection]
 [Date of requesting appeal against examiner's decision of rejection]
 [Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

BEST AVAILABLE COPY

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2004-170552

(P2004-170552A)

(43) 公開日 平成16年6月17日(2004.6.17)

(51) Int.C1.⁷

F I

| | | | | |
|--------|-------|--------|------|---------|
| G 10 L | 15/20 | G 10 L | 3/02 | 3 O 1 C |
| G 10 L | 11/02 | G 10 L | 3/00 | 5 1 3 C |
| G 10 L | 15/00 | G 10 L | 9/00 | 3 O 1 A |
| G 10 L | 15/02 | G 10 L | 3/00 | 5 5 1 A |
| G 10 L | 15/04 | G 10 L | 3/02 | 3 O 1 Z |

テーマコード(参考)

5 D 0 1 5

審査請求 未請求 請求項の数 5 O L (全 14 頁) 最終頁に続く

(21) 出願番号

特願2002-334276 (P2002-334276)

(22) 出願日

平成14年11月18日 (2002.11.18)

(71) 出願人 000005223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番
1号

(74) 代理人 100089118

弁理士 酒井 宏明

(72) 発明者 遠藤 香緒里

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 大田 恭士

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

(72) 発明者 佐々木 均

神奈川県川崎市中原区上小田中4丁目1番
1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】音声抽出装置

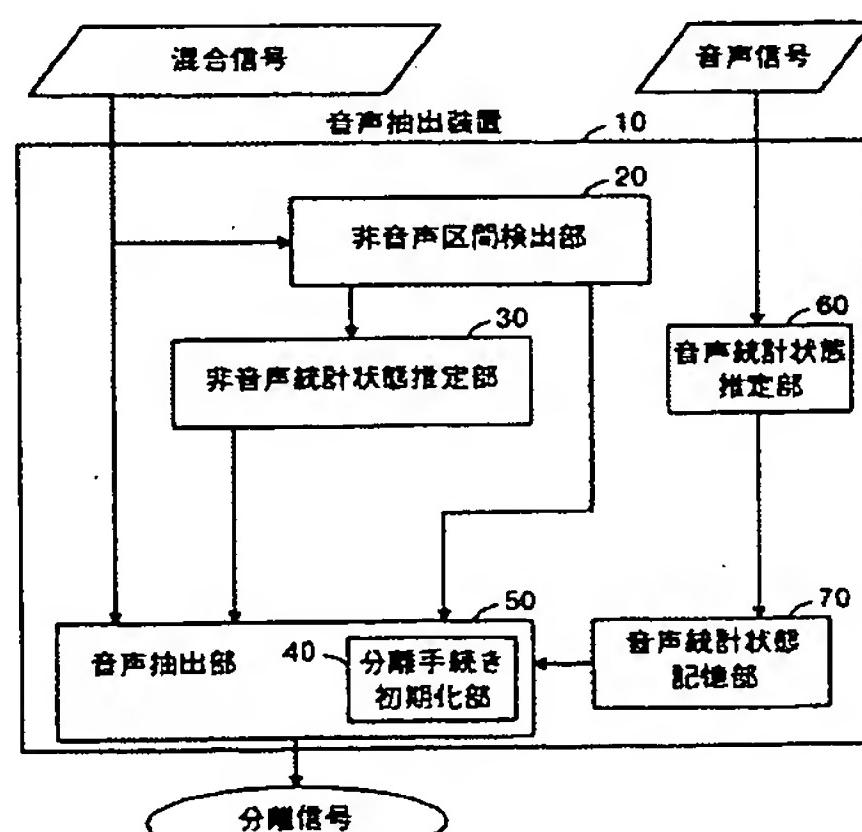
(57) 【要約】

【課題】特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、特定人の音声を、音源の時間に依存した統計状態を考慮しつつ、抽出すること。

【解決手段】非音声区間検出部20は、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出し、非音声統計状態推定部30は、検出された非音声区間の情報を用いて非音声の統計状態を推定し、推定された非音声の統計状態と音声統計状態記憶部70において知識として記憶されている音声に関する統計状態との統計的な独立性を用いて、音声抽出部50は、分離手続き初期化部40によって選択された初期手続きから、混合信号を逐次更新することによって少なくとも一つの音声を抽出する。

【選択図】 図2

音声抽出装置の機能ブロック図



【特許請求の範囲】**【請求項 1】**

特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出装置であって、
複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手段と、
前記非音声区間検出手段により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手段と、
非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手段と、
を備えたことを特徴とする音声抽出装置。

10

【請求項 2】

前記非音声区間検出手段は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする請求項 1 に記載の音声抽出装置。

【請求項 3】

前記非音声区間検出手段は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする請求項 1 に記載の音声抽出装置。

20

【請求項 4】

前記音声抽出手段は、前記非音声統計状態推定手段により推定された時間によって変動する非音声の統計状態と知識としての前記特定人の音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記特定人の音声の統計状態の組と前記混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする請求項 1 に記載の音声抽出装置。

【請求項 5】

前記非音声統計状態推定手段において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手段を備えたことを特徴とする請求項 1 に記載の音声抽出装置。

30

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明は、複数音源が混合した複数の入力信号と、該入力信号に含まれる音源の統計的な独立性を用いて、特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、特定人の音声を、音源の時間に依存した統計状態を考慮しつつ、抽出する音声抽出装置に関し、特に、携帯電話システムおよび実況中継において、混入した環境雑音と送話音声を分離し、送話音声のみを送信することができる音声抽出装置に関する。

40

【0002】**【従来の技術】**

複数の信号源から出た信号が混合している時、異なる信号源からでた信号が互いに統計的に独立であるという条件のみから、各々の信号を分離する方法は独立成分分析と呼ばれている。現在、よく知られている独立成分分析の手法（例えば、非特許文献 1 参照）では、信号源が未知で信号源の信号が一定の割合で線形的に結合され、観測されると仮定し、統計的な独立性の概念を基に混合信号を分離する手法を用いる技術が開示されている。

50

【0003】

また、独立成分分析において、混合信号の数が統計的に独立な信号源の個数と等しいか、あるいはそれ以上の場合、該独立信号の混合である混合信号を分離し、未知である信号源の信号を特定する方法は知られている。例えば、非特許文献 1 がある。しかしながら、非特許文献による手法では混合信号の数より多い統計的に独立な音源信号からなる混合信号を分離することはできない。

50

【0004】

これに対して、特許文献1は混合信号の数より多い独立な音源信号を混合信号の観測時間を遅延し、混合信号を逐次的に分離更新する方法を開示している。

【0005】**【非特許文献1】**

S. Amari, T. P. Chen, およびA. Cichocki共著論文「Stability analysis of learning algorithms for blind source separation」Neural Networks, Vol. 10, No. 8, pp. 1345-1351

【特許文献1】

特開2002-55969号公報

10

【0006】**【発明が解決しようとする課題】**

しかしながら、この特許文献1の技術では、混合信号を逐次的に分離更新する方法に時間に依存した統計的機構が含まれていないため、音源が移動している状況などを考えれば、信号源の混合のされ方自体が時間変化する状況では混合信号を適切に分離することが望めない。

【0007】

このため、混合信号の数以上の統計的に独立な音源信号からなる混合信号を、時間によって変動する混合信号の統計状態を考慮して、混合信号を逐次的に分離しつつ、目的に応じた音源信号を抽出することは、極めて重要な課題となっている。特に、携帯電話システムおよび実況中継においては、話者がマイクを持って移動する可能性があるため、環境雑音が話者の所有するマイクに混入する可能性がある。このような状況において、時間変化する環境雑音を考慮して、混入した環境雑音と送話音声を分離し、送話音声のみを送信する必要性がある。

20

【0008】

この発明は、上記従来技術による課題を解決するためになされたものであり、複数音源が混合した複数の入力信号と、該入力信号に含まれる音源の統計的な独立性を用いて、特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、特定人の音声を、音源の時間に依存した統計状態を考慮しつつ、抽出する音声抽出装置を提供することを目的とする。

30

【0009】**【課題を解決するための手段】**

本発明は、上記目的を達成するためになされたものであり、請求項1に係る音声抽出装置は、特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出装置であって、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手段（図2に示す非音声区間検出部20に対応する）と、前記非音声区間検出手段により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手段（図2に示す非音声統計状態推定部30に対応する）と、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手段（図2の音声抽出部50に対応する）と、を備えたことを特徴とする。

40

【0010】

この請求項1の発明によれば、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出し、検出された非音声区間の情報を用いて非音声の統計状態を推定し、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の声を抽出することとしたので、時間変化する環境雑音を伴う混合信号から目的とする音声信号を抽出することができる。

【0011】

また、請求項2に係る音声抽出装置は、請求項1の発明において、前記非音声区間検出手

50

段は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする。

【0012】

この請求項2の発明によれば、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することとしたので、単位時間毎に変化する環境雑音を非音声信号として抽出でき、該抽出結果を用いて、非音声信号の統計状態を推定することができる。

【0013】

また、請求項3に係る音声抽出装置は、請求項1の発明において、前記非音声区間検出手段は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする。10

【0014】

この請求項3の発明によれば、単位時間毎に混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうこととしたので、単位時間毎に混合信号の中から環境雑音を抽出し、単位時間毎に環境雑音の統計状態を推定することができる。

【0015】

また、請求項4に係る音声抽出装置は、請求項1の発明において、前記音声抽出手段は、前記非音声統計状態推定手段により推定された時間によって変動する非音声の統計状態と知識としての音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記音声の統計状態の組と混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする。20

【0016】

この請求項4の発明によれば、推定された時間によって変動する非音声の統計状態と知識としての音声に関する統計状態との統計的な独立性を用いて、非音声の統計状態と音声の統計状態の組と混合信号の統計状態との相互情報量の隔たりを、単位時間毎に逐次更新する分離手続きを行うこととしたので、混合信号の統計状態から音声信号の統計状態を分離でき、よって混合信号から目的とする音声信号を抽出することができる。

【0017】

また、請求項5に係る音声抽出装置は、請求項1の発明において、前記非音声統計状態推定手段において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手段（図2の分離手続き初期化部40に対応する）を備えたことを特徴とする。30

【0018】

この請求項5の発明によれば、混合信号の分離手続きを初期化することとしたので、手続きの初期状態から安定であり確度の高い音声信号の抽出ができる。

【0019】

【発明の実施の形態】

以下に添付図面を参照して、この発明に係る音声抽出装置の好適な実施の形態を詳細に説明する。なお、本実施の形態では、便宜上、2つの混合信号を分離する場合について説明するが、少なくとも一つの音声信号源並びに少なくとも一つの非音声信号源からなる混合信号を分離する場合についても、目的とする音声信号の抽出が可能である。40

【0020】

以下、本発明の概略構成を説明した後、音声抽出装置の処理手順の概念を説明する。最後に、後音声抽出装置の処理手順としていくつかの実施の形態をフローチャートに基づいて説明する。

【0021】

まず、本発明である音声抽出装置の概略構成を説明する。図1は本発明の概略構成を示す図である。複数の音源として、非音声音源1から非音声音源nまでのn個の音源と音声音源1から音声音源mまでのm個の音源から、各音声音源に対応する音声を抽出するための50

音声抽出装置1から音声抽出装置mまでがある。各音声抽出装置にはマイクが備わっておりここから非音声音源および音声音源から発信される混合信号を受信する。また、各音声抽出装置は、受信した混合信号から目的とする音声信号を抽出し、インターネットや無線あるいは有線の専用の通信回線を用いて、混合信号から分離された音声信号を別のサーバに出力する。

【0022】

次に、音声抽出装置の機能について説明する。図2は音声抽出装置の機能ブロック図である。同図に基づいて各機能を説明する。

【0023】

音声抽出装置10は、音声音源と非音声音源の信号からなる混合信号をオンラインで分離するために、予め各音声信号の統計状態を音声統計状態推定部60において推定し、推定結果を音声統計状態記憶部70に知識として記憶しておく。10

【0024】

非音声区間検出部20は、信号データのフレーム（信号源の数に対応した信号データをまとめたもので、実際は、ベクトル量として考える）毎に混合信号を入力し、該混合信号について当該フレームの音声または非音声の判定を行う。

【0025】

非音声統計状態推定部30は非音声区間検出部20より音声および非音声判定結果と混合信号を受け取り、非音声区間の場合に混合信号を用いて振幅値の頻度分布を作成する。統計状態は振幅値の頻度を観測数で割ることによって求めることができる。20

【0026】

分離手続き初期化部40は非音声区間検出部20より音声または非音声の判定結果と混合信号を、音声抽出部50より分離手続きの処理方法を受け取る。分離手続き初期化部40は非音声区間の混合信号を用いて、分離手続きの処理方法の初期手続きを決定する。音声抽出部50から入力した分離手続きの処理方法と分離手続き初期化部40によって決定された分離手続きの処理方法との間に、ある数量的に表された一定以上の違いがあれば、音声抽出部50の分離手続きの処理方法の初期化を行う。

【0027】

音声抽出部50は、混合信号、非音声統計状態推定部30から非音声信号の統計状態、および分離手続き初期化部40から初期化された分離手続きの処理方法との3つを受け取り、音声統計状態記憶部70に、知識として予め保持された音声信号源の統計状態と、非音声統計状態推定部30が算出した非音声音源の統計状態を用いて統計的独立性の尺度を算出し、独立性が高くなる規範で分離手続きの処理方法を更新する。該更新された分離手続きの処理方法を用いて音声信号と非音声信号に分離する。30

【0028】

以上の音声抽出装置の処理手続きの音声通信方式への応用例として、図3を説明する。同図に示すように、音声信号と非音声信号からなる混合信号は、音声抽出装置1の入力手段1と入力手段2によって、2つの混合信号として入力される。これら2つの混合信号は、音声抽出部50、100で分離され、音声信号および非音声信号となる。ここまで処理は音声抽出装置の処理であるが、更に音声抽出装置に音声符号化部110、送信部120を付加することによって、本発明による音声抽出方法を音声通信方式へ応用可能である。40

【0029】

以下、音声抽出装置の音声抽出部の処理説明に必要となる独立成分分析の具体的な説明を、数式を用いて説明する。独立な音源信号 s (s は音源信号ベクトル) がある時間定数である混合行列 A で混合され、混合信号 x (混合信号ベクトル) が観測されるものとする。数式で表現すると、 $x = As$ となる。この式を、以下、観測式と呼ぶことにする。

$$\text{音源信号ベクトル } s = (s_1(t), \dots, s_n(t))^T,$$

$$\text{混合信号ベクトル } x = (x_1(t), \dots, x_n(t))^T,$$

混合行列 A ($n \times n$ 行列)、

$$\text{分離結果の信号ベクトル } y = (y_1(t), \dots, y_n(t))^T$$

50

とする。

【0030】

このとき、混合信号 x を用いて信号の統計的な独立性を手掛りに分離行列 W を求め、音源を分離する。数式で表現すると、 $y = Wx$ となる。この式を、以下、分離式と呼ぶことにする。分離式で x を分離した（実際には、ベクトル x の左から行列 W を乗算した）際に、 y が独立となるように分離行列 W を求める問題となる。分離行列 W は音源間の統計的な独立性が大きくなる方向になるように下記の式で適応的に更新する。

【数1】

$$\Delta W = \eta [I - E[\phi(y)y^T]]W \quad \dots (1)$$

10

ここで、

【数2】

$$\phi(y) = -\left[\frac{\partial}{\partial y_1} \log P_1(y_1), \dots, \frac{\partial}{\partial y_n} \log P_n(y_n) \right]^T \quad \dots (2)$$

であり、 η は分離更新の大きさを調整する十分小さな正の係数である。

【0031】

(数1) は、分離結果の信号ベクトルの確率密度関数と分離結果の信号ベクトルの各周辺分布の積との間の相対エントロピー（カルバック・ライブラー情報量など）が計算され、分離行列 W の関数である該相対エントロピーが最小化される方向を勾配流の概念を用いて導出される。ここで、重要な事実は、分離結果の確率密度関数を算出する必要はなく、混合信号ベクトルの確率密度関数を導出すればよい。従って、分離結果のベクトルの確率密度関数をいかに混合信号ベクトルから推定すればよいかが問題となる。

20

【0032】

以下、音声抽出装置の処理手順に関する3つの実施の形態について説明する。以下では、混合信号の数を2 ($n = 2$)、分離行列 W (2×2 行列) の場合について具体的に説明する。

【0033】

(実施の形態1)

30

図4は、音声抽出装置の処理手順のフローチャート（その1）である。以下、各工程の説明を同図に基づいておこなう。各音源の確率密度関数を予め知識として記憶する（ステップS001）。ただし、環境雑音等、未知の音源の確率密度関数は当然ながら保持する必要はない。具体的には、例えば、携帯電話を所有する所有人物が、携帯電話を電話として使用する前に所有者の音声信号を携帯電話に記憶させ、その音声の振幅をもとに音声統計状態推定部60は確率密度関数を作成し、これを携帯電話の音声統計状態記憶部70に記憶しておく。

【0034】

次に、分離行列 W をランダム値で初期化する（ステップS002）。次に、1フレーム分のN個の混合信号を入力する（ステップS003）。例えば、 $x_1 = [x_1(0), \dots, x_1(k-1)]$, $x_2 = [x_2(0), \dots, x_2(k-1)]$, ここで、 k は1フレームのサンプル数をあらわす。次に、式 $y = Wx$ により、分離行列 W を用いて音源としての混合信号ベクトル x を分離する（ステップS004）。式 $y = Wx$ を成分で書くと

40

【数3】

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \quad \dots (3)$$

となる。

50

【0035】

次に、分離信号のうち非音声信号源を特定する（ステップS005）。実際、式 $y = Wx$ で求めた分離信号 y の成分のうち、非音声に対応する成分を選ぶ。 y の成分のピッチ相関を算出し、最もピッチ相関が小さいものを非音声に対応する成分とする。非音声に対応する成分が y_1 成分となるように、 y 、 W を並べ替える。たとえば、 y_2 成分が非音声に対応する成分の場合には、 y_1 、 y_2 を入れ替える。これに伴い W の成分も第一行と第二行を入れ替える。次に、混合信号1の音声または非音声判定を行う。例えば、ピッチ相関を算出し、ピッチ相関く閾値の場合、非音声と判定し、ピッチ相関 \geq 閾値の場合、音声と判定する（ステップS006）。

【0036】

次に、混合信号1が非音声と判定された場合は、ステップS008に処理を進める。混合信号1が音声と判定された場合は、ステップS013に処理を進める（ステップS007）。次に、混合信号2の音声および非音声判定を行う（ステップS008）。例えば、ピッチ相関を算出し、ピッチ相関く閾値の場合、非音声と判定し、ピッチ相関 \geq 閾値の場合、音声と判定する。次に、混合信号2が非音声と判定された場合は、ステップS010に処理を進め、混合信号2が音声と判定された場合は、ステップS013に処理を進める（ステップS009）。

【0037】

次に、混合信号1、2をパワーで正規化する（ステップS010）。実際、 $x_1(t) = x_1'(t) / p_1$ および $x_2(t) = x_2'(t) / p_2$ ($t = 0, \dots, k-1$)。 p_1 、 p_2 をそれぞれ y_1 、 y_2 のフレームパワーとする。非音声信号源の確率密度関数を算出する（ステップS011）。 $x_1'(t)$ および $x_2'(t)$ を用いて、正規化した非音声区間の信号の振幅の頻度分布を作成する。 $x_1'(t)$ および $x_2'(t)$ は、-1から+1の範囲の値となるので、-1から+1の範囲を適当な数Rに分割し、 $r(i) = q(i) / N_{a,i}$ として非音声信号源の確率密度関数を算出する。ここで、 i は $0, \dots, R-1$ の範囲を動き、 $r(i)$ は非音声信号源の確率密度関数、 $q(i)$ は i 番目の区間にに入る振幅値の頻度（起動時からの総数）、 $N_{a,i}$ は振幅値の数（起動時からの総数）である。

【0038】

次に、非音声信号源の確率密度関数をステップS011で算出したもので置き換える（ステップS012）。次に、分離行列Wを（数1）によって更新する（ステップS013）。次に、ステップS004で算出した分離信号を出力する（ステップS014）。次に、入力終了の場合は、処理を終了し、入力が続く場合は、ステップS003に処理を進める（ステップS015）。

【0039】

上述してきたように、本実施の形態1では、非音声区間検出部20で入力信号の非音声区間を検出することで、非音声のみの信号を知ることができ、これを用いて非音声信号源の確率密度関数を算出できる。これにより非音声信号源が未知である場合や時間変化する場合にも確率密度関数を正確に求めることができ、したがって分離性能の向上が可能となる。

【0040】**(実施の形態2)**

図5は、音声抽出装置の処理の処理手順のフローチャート（その2）である。以下、各工程の説明を同図に基づいておこなう。ステップS101～ステップS109までは、実施の形態1と全く同様であるから、説明を省略する。ステップS110以降を説明する前に、本実施の形態の本質部分である分離行列Wの初期値算出の具体的方法について説明をした後、各ステップの説明を行う。

【0041】

分離行列Wの初期値の算出は下記の通りに行う。時刻 t で音源 $s_1(t)$ と $s_2(t)$ が混合行列Aで混合され、その混合結果 $x_1(t)$ と $x_2(t)$ が観測されるものとする。

10

20

30

40

50

音源 s_1 は音声信号源、音源 s_2 は非音声信号源であるとする。また混合行列は時間で変化しないと仮定する。数式で表現すれば下記の（数4）となる。

【数4】

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} \quad \cdots (4)$$

$x_1(t)$ と $x_2(t)$ を観測し、元の音源に分離するような分離行列 W を推定することで、音源を分離する。数式で表現すれば（数3）となる。

【0042】

ここで、音声検出手段により、非音声区間を判定すると、その区間では観測される信号には非音声しか含まれていないことと、 $A = W^{-1}$ より、 $A_{1,2} / A_{2,2} = -W_{1,2} / W_{2,2} = \alpha$ となる。したがって、 W の初期値を X とすると、 X の第一行の成分が、 $X_{1,2} / X_{2,2} = \alpha$ を満たすように W の初期値 X を算出する。ただし、 X の第二行目の成分は自由変数である。

【0043】

さて、ここで処理手順のステップの説明に戻る。上記説明に基づき、分離行列の初期値 X を算出する（ステップS110）。次に、現在の分離行列 W と初期値 X について $| - (W_{1,2} / W_{2,2}) - (X_{1,2} / X_{2,2}) | > TH$ を評価し、成り立つ場合は、ステップS112に処理を進め、成り立たない場合は、ステップS113に処理を進める（ステップS111）。次に、分離行列を次式で更新する（ステップS112）。 $W_{1,1} = W_{1,1}$, $W_{1,2} = (X_{1,2} / X_{2,2}) \times W_{1,1}$, $W_{2,1} = W_{2,1}$, $W_{2,2} = W_{2,2}$ 。次に、ステップS4で算出した分離信号を出力する（ステップS113）。分離行列 W を（数1）によつて更新する（ステップS114）。次に、ステップS104で算出した分離信号を出力する（ステップS115）。次に、入力終了の場合は、処理を終了し、入力が続く場合は、ステップS103に処理を進める（ステップS116）。

【0044】

上述してきたように、本実施の形態2では、非音声区間検出部20で入力信号の非音声区間を検出することで、非音声のみの信号を知ることができる。これを用いて分離行列の初期値を算出できる。このことでランダムな初期値よりも正解に近い初期値から計算を出発できること、また環境騒音の変化により分離行列が正解から外れた場合に、正解に近い初期値に設定できるので、正しい分離行列を求め易くなる。

【0045】

（実施の形態3）

図6は、音声抽出装置の処理手順のフローチャート（その3）である。以下、各工程の説明を同図に基づいておこなう。実施の形態3は、実施の形態1および実施の形態2を組み合わせたものであり、混合信号をパワーで正規化する処理と分離行列の初期値を算出する処理を音声抽出処理に組み込んだものである。

【0046】

以下、各工程の説明を同図に基づいておこなう。各音源の確率密度関数を予め知識として記憶する（ステップS201）。ただし、環境雑音等、未知の音源の確率密度関数は当然ながら保持する必要はない。具体的には、例えば、携帯電話を所有する所有人物が、携帯電話を電話として使用する前に所有者の音声信号を携帯電話に記憶させ、その音声の振幅をもとに確率密度関数を作成し、これを携帯電話の記憶部に記憶しておく。次に、分離行列 W をランダム値で初期化する（ステップS202）。

【0047】

次に、1フレーム分のN個の混合信号を入力する（ステップS203）。例えば、 $x_1 = [x_1(0), \dots, x_1(k-1)]$, $x_2 = [x_2(0), \dots, x_2(k-1)]$ ，ここで、 k は1フレームのサンプル数をあらわす。次に、式 $y = Wx$ により、分離行列 W を用いて音源としての混合信号ベクトル x を分離する（ステップS204）。式 y

10

20

30

40

50

=W x を成分で書くと

【数3】

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} \quad \dots (3)$$

となる。

【0048】

次に、分離信号のうち非音声信号源を特定する（ステップS205）。実際、式 $y = Wx$ で求めた分離信号 y の成分のうち、非音声に対応する成分を選ぶ。 y の成分のピッチ相関を算出し、最もピッチ相関が小さいものを非音声に対応する成分とする。非音声に対応する成分が y_1 成分となるように、 y 、 W を並べ替える。たとえば、 y_2 成分が非音声に対応する成分の場合には、 y_1 、 y_2 を入れ替える。これに伴い W の成分も第一行と第二行を入れ替える。次に、混合信号1の音声または非音声判定を行う。例えば、ピッチ相関を算出し、ピッチ相関<閾値の場合、非音声と判定し、ピッチ相関≥閾値の場合、音声と判定する（ステップS206）。

10

【0049】

次に、混合信号1が非音声と判定された場合は、ステップS208に処理を進める。混合信号1が音声と判定された場合は、ステップS216に処理を進める（ステップS207）。次に、混合信号2の音声および非音声判定を行う（ステップS208）。例えば、ピッチ相関を算出し、ピッチ相関<閾値の場合、非音声と判定し、ピッチ相関≥閾値の場合、音声と判定する。次に、混合信号2が非音声と判定された場合は、ステップS210に処理を進め、混合信号2が音声と判定された場合は、ステップS216に処理を進める（ステップS209）。

20

【0050】

次に、混合信号1、2をパワーで正規化する（ステップS210）。実際、 $x_1(t) = x_1'(t) / p_1$ および $x_2(t) = x_2'(t) / p_2$ 、($t = 0, \dots, k-1$)。 p_1 、 p_2 をそれぞれ y_1 、 y_2 のフレームパワーとする。非音声信号源の確率密度関数を算出する（ステップS211）。 $x_1'(t)$ および $x_2'(t)$ を用いて、正規化した非音声区間の信号の振幅の頻度分布を作成する。 $x_1'(t)$ および $x_2'(t)$ は、-1から+1の範囲の値となるので、-1から+1の範囲を適当な数 R に分割し、 $r(i) = q(i) / N_{a,1}$ として非音声信号源の確率密度関数を算出する。ここで、 i は $0, \dots, R-1$ の範囲を動き、 $r(i)$ は非音声信号源の確率密度関数、 $q(i)$ は i 番目の区間にに入る振幅値の頻度（起動時からの総数）、 $N_{a,1}$ は振幅値の数（起動時からの総数）である。

30

【0051】

次に、非音声信号源の確率密度関数をステップS211で算出したもので置き換える（ステップS212）。次に、分離行列の初期値 X を算出する（ステップS213）。次に、現在の分離行列 W と初期値 X について $| - (W_{1,2} / W_{1,1}) - (X_{1,2} / X_{1,1}) | > TH$ を評価し、成り立つ場合は、ステップS215に処理を進め、成り立たない場合は、ステップS216に処理を進める（ステップS214）。次に、分離行列を次式で更新する（ステップS215）。 $W_{1,1} = W_{1,1}$ 、 $W_{1,2} = (X_{1,2} / X_{1,1}) \times W_{1,1}$ 、 $W_{2,1} = W_{2,1}$ 、 $W_{2,2} = W_{2,2}$ 。

40

【0052】

分離行列 W を（数1）によって更新する（ステップS216）。次に、ステップS204で算出した分離信号を出力する（ステップS217）。次に、入力終了の場合は、処理を終了し、入力が続く場合は、ステップS203に処理を進める（ステップS218）。

【0053】

上述してきたように、本実施の形態3では、非音声区間検出部20で入力信号の非音声区間を検出することで、非音声のみの信号を知ることができるので、これを用いて非音声統

50

計状態推定部30で非音声信号源の確率密度関数と分離手続き初期化部40で分離行列の初期値を算出できる。このことで非音声信号源が未知である場合や時間変化する場合にも確率密度関数を正確に求めることができる。さらに、ランダムな初期値よりも正解に近い初期値から計算を出発でき、また環境騒音の変化により分離行列が正解から外れた場合に、正解に近い初期値に設定できるので、正しい分離行列を求め易くなる。これらにより、分離性能の向上が可能となる。

【0054】

ところで、本実施の形態では、音声抽出装置の処理手順に関する3つの実施の形態について、混合信号の数を2($n=2$)、分離行列W(2×2行列)の場合について具体的に説明したが、本発明はこれに限定されるものではなく、nが一般の場合においても本発明を適用することもできる。10

【0055】

(付記1) 特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出装置であって、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手段と、前記非音声区間検出手段により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手段と、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手段と、20
を備えたことを特徴とする音声抽出装置。

【0056】

(付記2) 前記非音声区間検出手段は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特微量を判定することにより非音声区間を検出することを特徴とする付記1に記載の音声抽出装置。

【0057】

(付記3) 前記非音声区間検出手段は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする付記1に記載の音声抽出装置。30

【0058】

(付記4) 前記音声抽出手段は、前記非音声統計状態推定手段により推定された時間によって変動する非音声の統計状態と知識としての前記特定人の音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記特定人の音声の統計状態の組と前記混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする付記1に記載の音声抽出装置。

【0059】

(付記5) 前記非音声統計状態推定手段において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手段を備えたことを特徴とする付記1に記載の音声抽出装置。40

【0060】

(付記6) 特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出プログラムであって、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間検出手順と、前記非音声区間検出手順により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定手順と、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出手順と、
を備えたことを特徴とする音声抽出プログラム。50

【0061】

(付記7) 前記非音声区間検出手順は、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することにより非音声区間を検出することを特徴とする付記6に記載の音声抽出プログラム。

【0062】

(付記8) 前記非音声区間検出手順は、単位時間毎に前記混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうことにより非音声区間を検出することを特徴とする付記6に記載の音声抽出プログラム。

【0063】

(付記9) 前記音声抽出手順は、前記非音声統計状態推定手順により推定された時間によって変動する非音声の統計状態と知識としての前記特定人の音声の統計状態との統計的な独立性を用いて、前記非音声の統計状態と前記特定人の音声の統計状態の組と前記混合信号の統計状態との相互情報量の隔たりを、前記単位時間毎に逐次更新する分離手続きを行って、前記特定人の音声を抽出することを特徴とする付記6に記載の音声抽出プログラム。

【0064】

(付記10) 前記非音声統計状態推定手順において算出された統計状態を用いて、前記混合信号の前記分離手続きを初期化する分離手続き初期化手順を備えたことを特徴とする付記6に記載の音声抽出プログラム。

【0065】

(付記11) 特定人の音声信号と該音声信号以外の非音声信号とが混合された混合信号から、前記特定人の音声信号を抽出する音声抽出方法であって、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出する非音声区間声検出工程と、前記非音声区間検出工程により検出された非音声区間の情報を用いて非音声の統計状態を推定する非音声統計状態推定工程と、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の音声信号を抽出する音声信号抽出工程と、を備えたことを特徴とする音声抽出方法。

【0066】

【発明の効果】

以上説明したように、請求項1の発明によれば、複数の音源が混合した混合信号から、音声が含まれない非音声区間を検出し、検出された非音声区間の情報を用いて非音声の統計状態を推定し、非音声の統計状態と特定人の音声の統計状態とを用いて、前記特定人の声を抽出することとしたので、時間変化する環境雑音を伴う混合信号から目的とする音声信号を抽出することが可能な音声抽出装置が得られるという効果を奏する。

【0067】

また、請求項2の発明によれば、単位時間毎に前記混合信号の情報を用いて、音声と非音声との信号特徴量を判定することとしたので、単位時間毎に変化する環境雑音を非音声信号として抽出でき、該抽出結果を用いて、非音声信号の統計状態を推定することが可能な音声抽出装置が得られるという効果を奏する。

【0068】

また、請求項3の発明によれば、単位時間毎に混合信号間のピッチ相関を算出し、該ピッチ相関を用いて、混合信号が音声信号または非音声信号であるかの判定をおこなうこととしたので、単位時間毎に混合信号の中から環境雑音を抽出し、単位時間毎に環境雑音の統計状態を推定することが可能な音声抽出装置が得られるという効果を奏する。

【0069】

また、請求項4の発明によれば、推定された時間によって変動する非音声の統計状態と知識としての音声に関する統計状態との統計的な独立性を用いて、非音声の統計状態と音声の統計状態の組と混合信号の統計状態との相互情報量の隔たりを、単位時間毎に逐次更新

10

20

30

40

50

する分離手続きを行うこととしたので、混合信号の統計状態から音声信号の統計状態を分離でき、よって混合信号から目的とする音声信号を抽出することが可能な音声抽出装置が得られるという効果を奏する。

【0070】

また、請求項5の発明によれば、混合信号の分離手続きを初期化することとしたので、手続きの初期状態から安定であり確度の高い音声信号の抽出が可能な音声抽出装置が得られるという効果を奏する。

【図面の簡単な説明】

- 【図1】本発明の概略構成を示す図である。
- 【図2】音声抽出装置の機能ブロック図である。
- 【図3】本発明の音声通信方式への適用例の図である。
- 【図4】音声抽出装置の処理手順のフローチャート（その1）である。
- 【図5】音声抽出装置の処理手順のフローチャート（その2）である。
- 【図6】音声抽出装置の処理手順のフローチャート（その3）である。

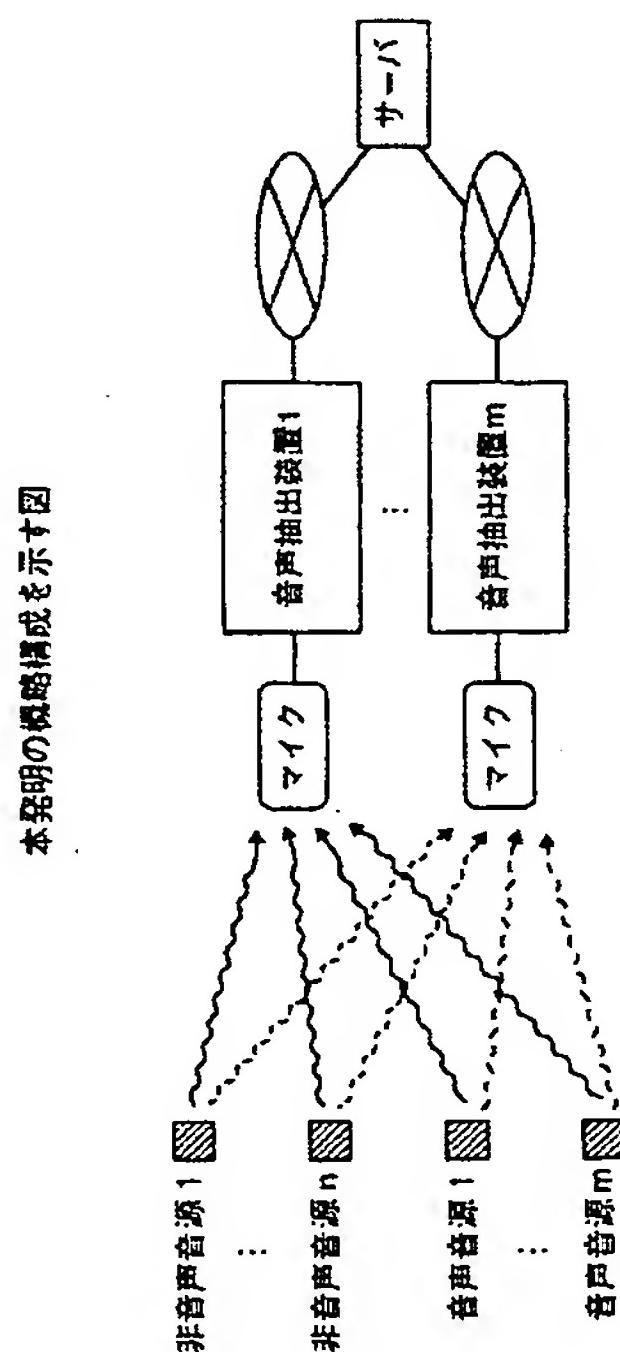
【符号の説明】

- 10 音声抽出装置
- 20 非音声区間検出部
- 30 非音声統計状態推定部
- 40 分離手続き初期化部
- 50, 100 音声抽出部
- 60 音声統計状態推定部
- 70 音声統計状態記憶部
- 110 音声符号化部
- 120 送信部

10

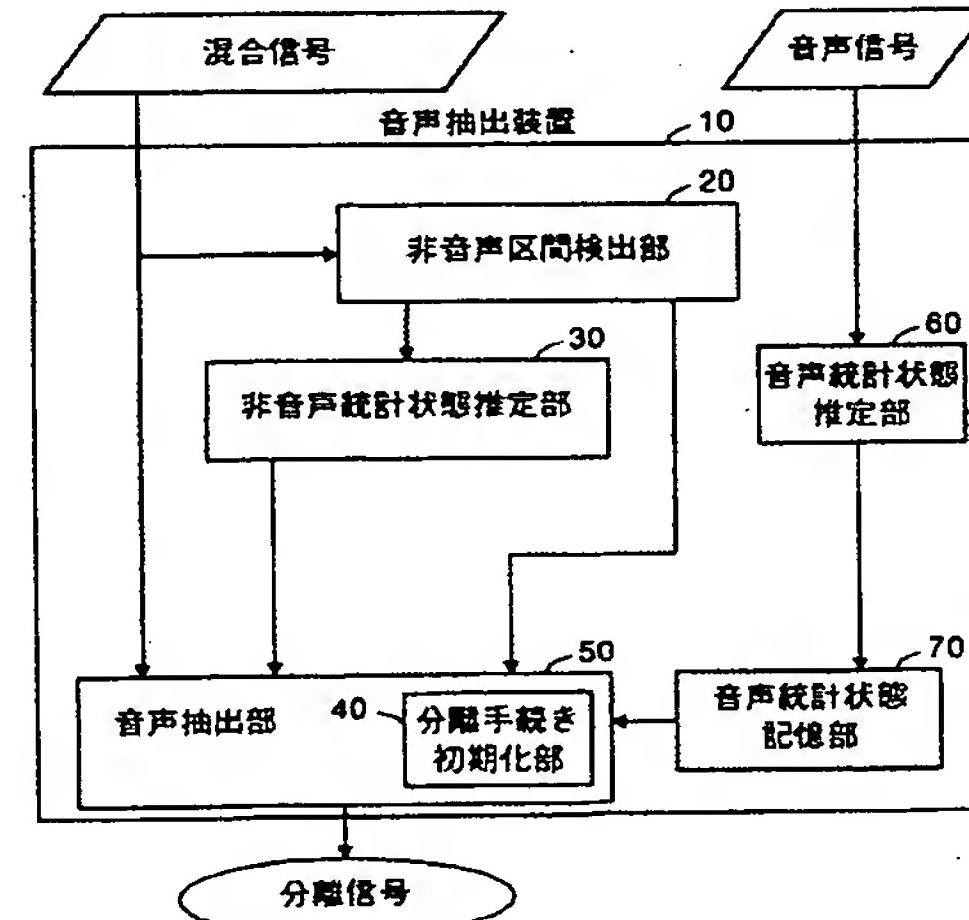
20

【図1】

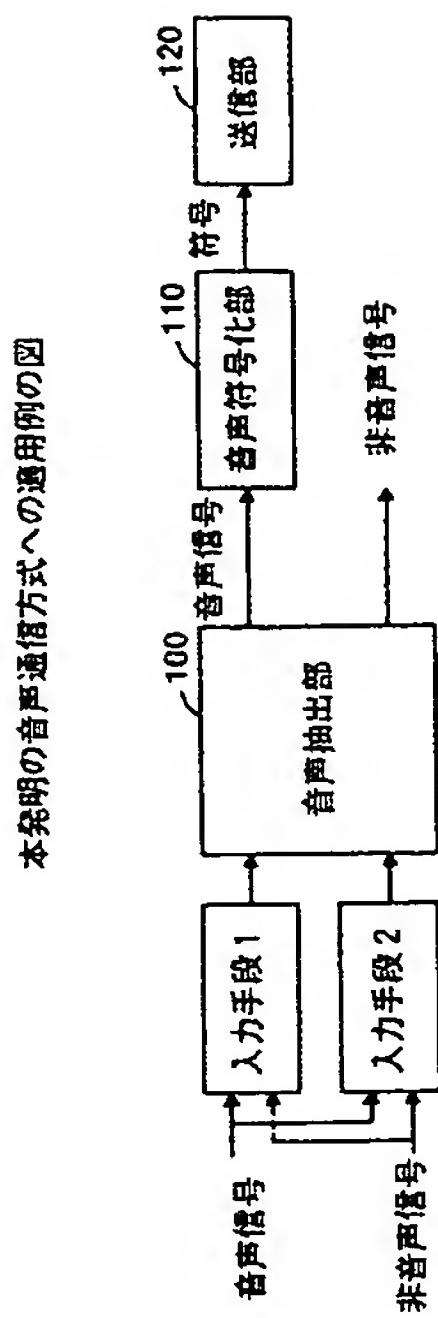


【図2】

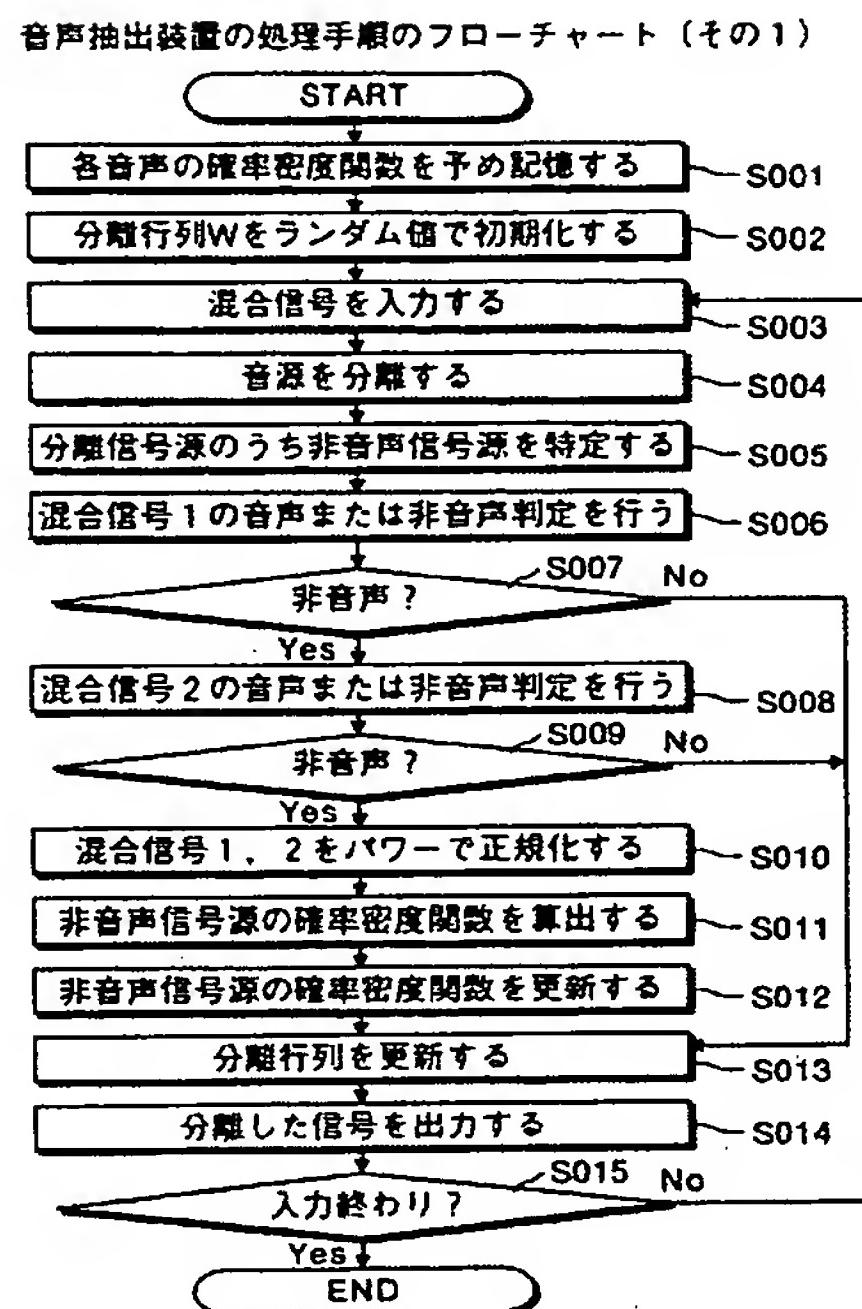
音声抽出装置の機能ブロック図



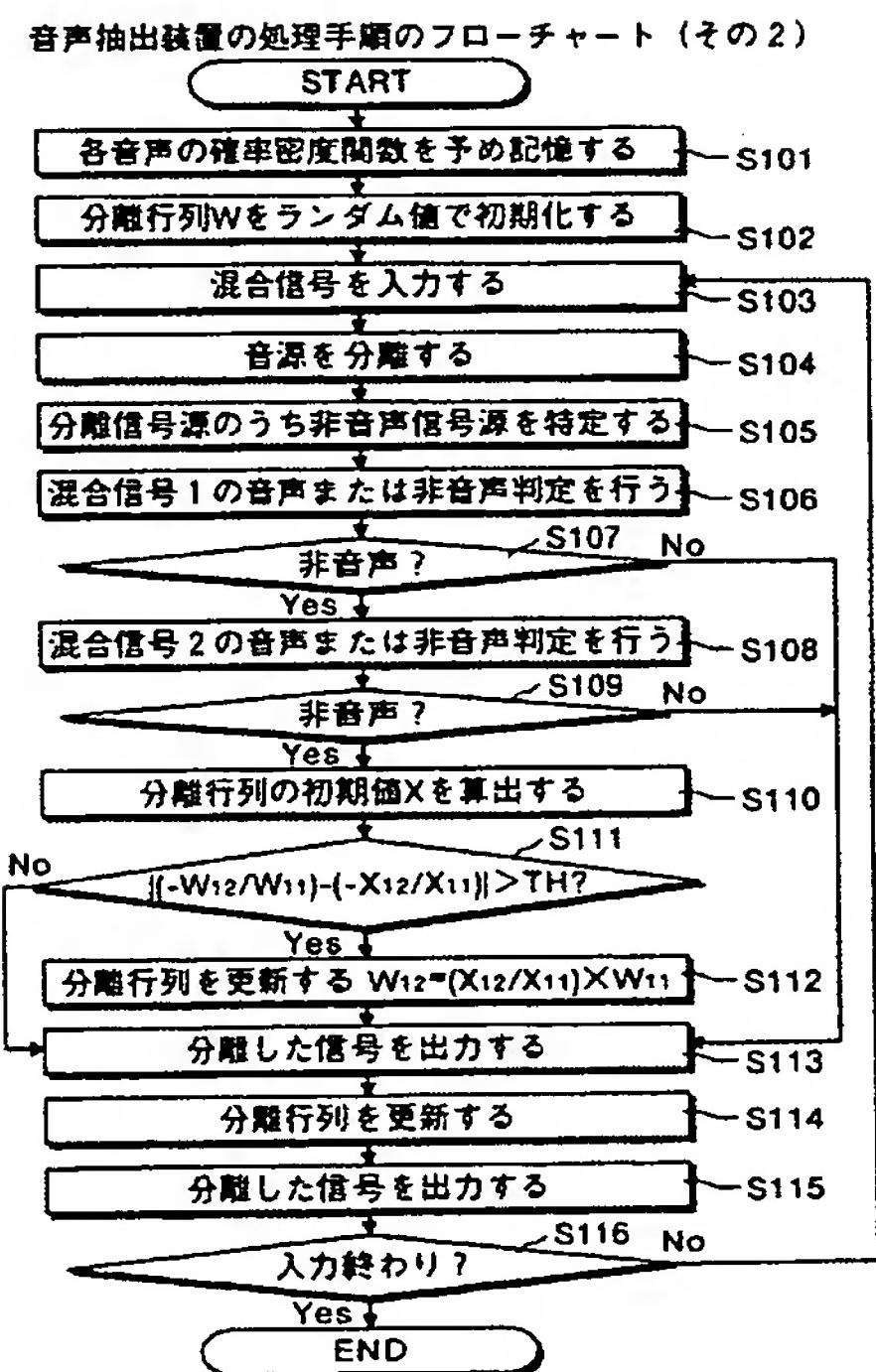
【図3】



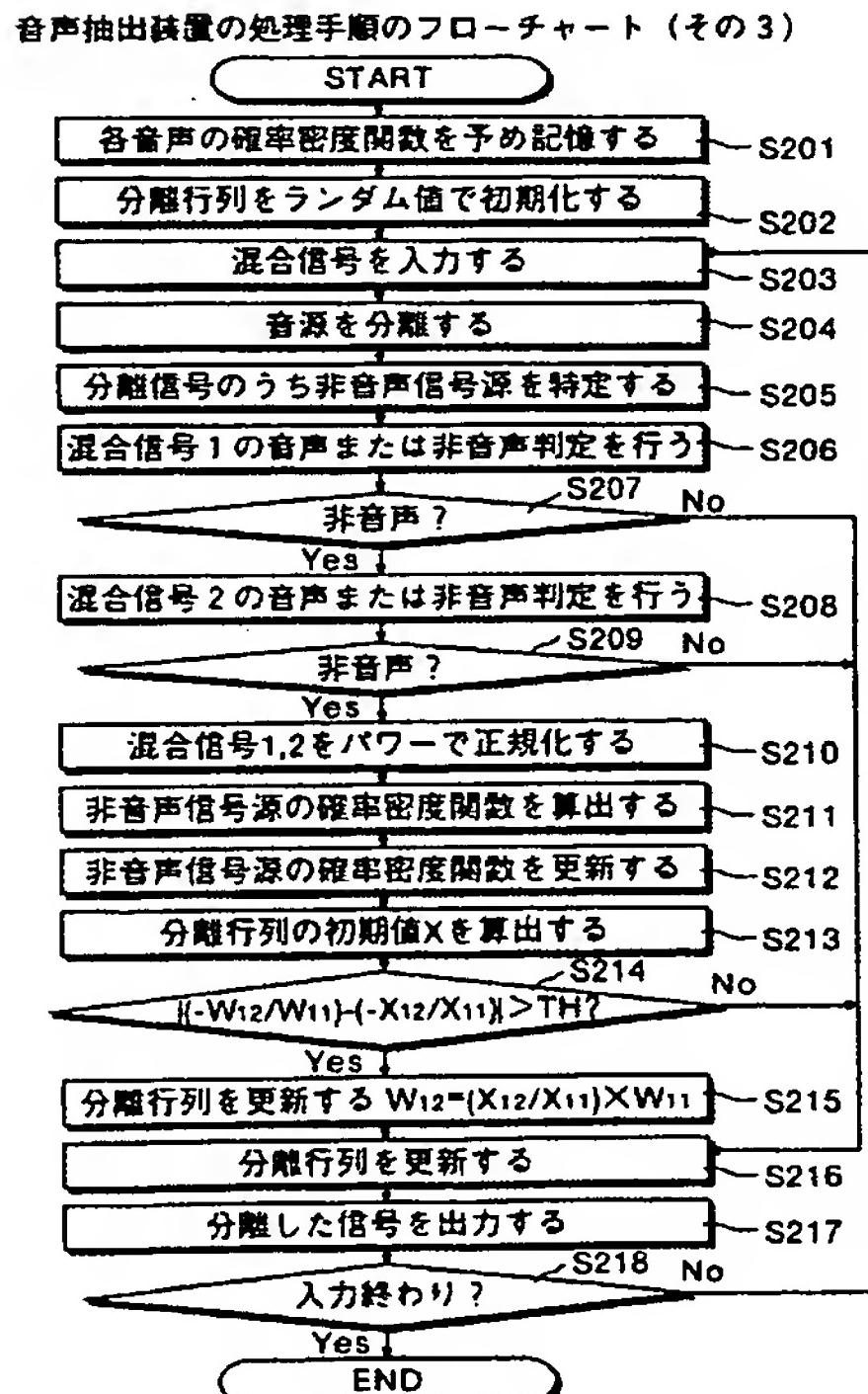
【図4】



【図5】



【図6】



フロントページの続き

(51)Int.Cl.'

F I

テーマコード(参考)

G 10 L 21/02

(72)発明者 松原 光良

福岡県福岡市博多区博多駅前三丁目22番8号 富士通九州ディジタル・テクノロジ株式会社内

F ターム(参考) 5D015 AA03 DD03 EE04